An Empirical Study of Epidemic Algorithms in Large Scale Multihop Wireless Networks

Deepak Ganesan, Bhaskar Krishnamachari, Alec Woo, David Culler, Deborah Estrin, and Stephen Wicker

IRB-TR-02-003

March, 2002

Intel Research Berkeley

An Empirical Study of Epidemic Algorithms in Large Scale Multihop Wireless Networks

Abstract A new class of networked systems is emerging that involve very large numbers of small, low-power, wireless devices. We present findings from a large scale empirical study involving over 150 such nodes operated at various transmission power settings. The study reveals that even a simple epidemic protocol, flooding, can exhibit surprising complexity at scale. The instrumentation in our experiments permits us to separate effects at the various layers of the protocol stack. At the physical/link layer, we present statistics on packet loss, effective communication range and link asymmetry; at the MAC layer, we measure contention, collision and latency; and at the network/application layer, we analyze the structure of trees constructed using flooding. The data and analysis lay a foundation for a much wider set of algorithmic studies in this space.

1 Introduction

A new class of networked systems is emerging that involve very large numbers of small, low-power, wireless devices distributed over physical space. Today, numerous investigations in wireless sensor networks are utilizing hundreds of battery powered nodes that are perhaps a cubic inch in size [25, 34]. Laboratory studies have demonstrated nodes of a few cubic millimeters in volume, and we can certainly imagine many scenarios of computational fabrics, surfaces, and floating dust [6, 8]. The sheer number of devices involved in such networks and the resource constraints of the nodes – energy, storage, and processing – motivate us to explore extremely simple algorithms for discovery, routing, multicast, and aggregation. These algorithms should be localized, use minimal state, adapt to changes in structure, and have minimal communication cost. Our experience has been that these algorithms, while easy to build, often exhibit complex global behavior in real world settings. In particular, phenomena that may be side issues in typical wireless LAN environments become quite significant. In this paper, we provide a wealth of detailed empirical data from studies of relatively large scale wireless network configurations to serve as a basis for algorithm design in this emerging space. We utilize this data to analyze the complex behavior exhibited by a simple epidemic multicast mechanism. In particular, we isolate the various factors influencing the global behavior by separating out each primary level: the physical/link layer, medium access layer, and network/application layer. The data and analysis lay a foundation for a much wider set of algorithmic studies.

The paper is organized as follows: in Section 2 we outline epidemic algorithms and show that even a simple epidemic protocol, flooding, exhibits surprising complexity at scale. We highlight some of the related work in Section 3. We discuss the experimental platform in Section 4 and describe our experiments, our methodology for examining the results and the need for useful metrics in Section 5. This sets up our discussion of the experimental results showing physical and link layer effects (Section 6), medium access layer effects (Section 7), and network/application layer effects (Section 8). We discuss the implications of this epidemiological study for the design of protocols for large scale wireless networks in Section 9.

2 Epidemic Algorithms: Motivating Scenario

We use the phrase *epidemic algorithms* to refer to network protocols that allow rapid dissemination of information from a source through purely local interactions. In an epidemic algorithm, a message initiated from a source is rebroadcasted by neighboring nodes and extends outward, hop by hop, until the entire network is reached. Figure 1 shows the schema for message handling in a generalized epidemic protocol. As seen in this schema, in the most general case, each node that receives a message may be induced by the content of the message to take some local action. It then makes a decision on whether or not

it should re-transmit a message. This retransmitted message could be the same as the original message or some modified version thereof. *Flooding*, in which nodes always re-transmit the message upon first reception, is a simple example of an epidemic algorithm. Sophisticated forms of flooding include probabilistic, counter-based, distance-based, location-based, and cluster-based flooding techniques [2], [20], [21], in which the retransmissions of packets are inhibited in order to minimize redundancy.

```
Let S be local state of node and R a random number.
If message M_i is received for the first time, then
  Take local action based on M_i:  S ← f_1(M_i, S).
  Compose message M_i' = f_2(M_i, S).
  Make Boolean retransmit decision D = f_3(S, R).
  If D is yes, then
    Transmit M_i' to all neighbors.
```

Figure 1: Schema for message handling in a generic epidemic protocol

Such epidemic algorithms underly more sophisticated protocols, particularly in large scale multihop wireless networks, in which there may be a need for unattended operation. For example, they are used for single source-destination route discovery in reactive and hybrid ad hoc routing protocols [10], [11], [12]; for exploration in directed diffusion [43]; for multi-hop broadcast [2], [17]; for forming discovery trees [22]; for issuing network commands such as "sleep," "wake up"; for changing network-wide parameters such as transmit power, and for multihop time synchronization [27].

In this study we focus on a simple epidemic protocol: flooding. Under idealized settings, one would expect a flood to ripple outward from the source in an orderly, uniform, fashion. Our results from a large-scale implementation of this protocol on real hardware show, however, that the global behavior of this simple protocol can be surprisingly complex. We now present a motivating example to support this point.

Figure 2 shows a sequence of snapshots from traces of an experiment to illustrate how a flood propagated over time[1]. 160 nodes are laid out in a 12x13 square grid on the ground as indicated by dots and a flood originates from the node located at the coordinates (5,0). When a node receives the flood message, it immediately rebroadcasts once and squelches further retransmissions. Redundancy is expected as every node responds in this manner.

There are several noteworthy indicators of non-uniform flood propagation. Instead of extending outward step-by-step, there were links that show the flood actually extends backward geographically towards the source. We call these *backward links*. An example of this can be seen in Figure 2(b) - the link between the node located at (6,3) and the node located at (6,2) is a backward link; Figure 2(c) also shows numerous backward links. In some instances, a flooding message is received over a large distance and creates what we call a *long link*, such as the links from the source near (5,0) to nodes at (1,1) and (2,3) in Figure 2(a). Finally, some nodes are missed by the flood even though neighboring nodes transmit messages, such as the node near (3,4) in Figure 2(c). We refer to these nodes as *stragglers*. If we look at the tree structure that evolves as the flooding proceeds, we notice that it exhibits a high *clustering* behavior: most nodes in the tree have few or no descendants, while a significant few have many children.

This simple example illustrates how such epidemic processes can exhibit complex behavior in a realistic setting. There are a number of factors across different layers that impact the dynamics of flooding. For example, the long links show that the cell region for each node is far from a simple disc (a physical/link level effect), and the stragglers are very likely left behind due to MAC-level collisions. In the following sections, we dissect the contributions of different layers to the behavior of flooding at scale. Our experimental observations provide valuable input into the design of more sophisticated multicast mechanisms for large scale wireless networks.

3 Related Work

There is currently a dearth of experimental measurements on the scaling of ad-hoc protocols. Prior experimental studies in this area have tended to focus on routing in wireless ad-hoc networks without addressing scaling for lack of large enough infrastructure. For example, [3] describes an experimental ad-hoc network with eight mobile nodes consisting of laptops and 802.11 cards driven around in a 300 x 700 m area. This test-bed is used to provide some results on the performance of dynamic source routing (DSR). Similarly, [4] describes an experimental testbed involving one desktop and five laptops with 802.11 cards, used to test the performance of the ad hoc on-demand distance vector (AODV) routing protocol. In [28], the performance of data aggregation in directed diffusion is tested on a sensor network consisting of 14 PC/104 nodes equipped with Radiometrix RPC modems. Some experimental work has sought to validate and test medium access protocols. An 11-node experi-

[1]The unmarked nodes in Figure 2 correspond to failed nodes in the experiment

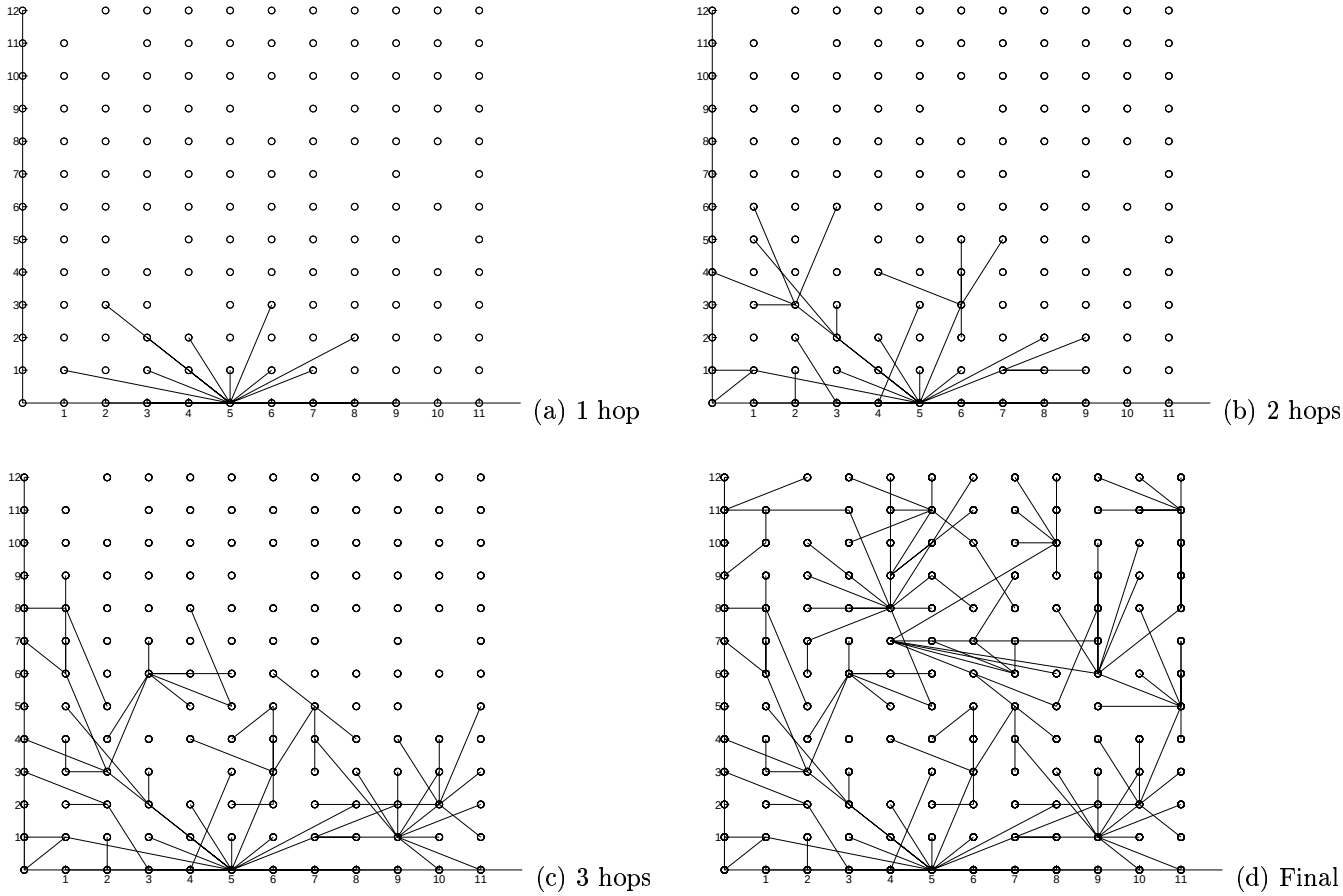


Figure 2: Snapshots from a single run of flooding on the experimental testbed

mental setup using Berkeley motes is used to analyze the performance of an adaptive rate control mechanism for medium access in sensor networks in [30]. In [29], a small experimental setup consisting of 5 Berkeley motes is used to validate the performance of the proposed S-MAC protocol. Another small-scale experiment involving Berkeley motes for signal strength measurements is described in [31].

Most previous work on analyzing the behavior of routing protocols in large-scale multihop wireless networks has been done in simulation [32][43]. These studies are not entirely satisfactory because the realistic modelling of physical and link-layer characteristics in simulation settings is a very challenging problem, and the final validation of the protocol performance has to be in real settings.

We note that the phrase *epidemic algorithms* has been widely used in the context of replicated databases [5]. In this paper we have chosen to focus on a simple protocol for epidemic dissemination of information through a wireless network: flooding. This simple mechanism is also the subject of [2] which investigates the "broadcast storm" problem associated with flooding, showing both analytically and through simulations inefficiencies such redundancy, MAC-level contention and collision. To a large extent, our work in this paper provides real experimental support to the results in [2]. But, we go a step further in our work by examining the impact of physical/link layer non-idealities as well on the dynamic behavior of this simple epidemic protocol.

The authors in [2] also propose five schemes which alleviate the problem by inhibiting some nodes from rebroadcasting: probabilistic (which is similar to Gossip [20], [21]), counter-based, distance-based, location-based, and cluster-based. It can be seen that these schemes are essentially more sophisticated epidemic protocols as they fit the schema shown in Figure 1.

While epidemic algorithms have some advantages in the face of high mobility [17], they may incur relatively high redundancy. There are other mechanisms for broadcast information dissemination in multihop wireless networks. The SPIN protocols presented in [36] efficiently disseminate information among sensors in an energy-constrained

multihop wireless sensor network by using meta-data negotiations to eliminate redundant signalling. A reliable broadcast service for wireless networks based on a multicluster architecture that is more efficient than flooding is described in [19]. There is also a line of work which has considered protocols involving the construction of a efficient minimum connected dominating sets within the network in order to perform broadcasts in wireless networks with optimal signalling costs [14], [15], [16].

4 Experimental Platform

The nodes used in these experiments are shown in Figure 4. Each node has a 4 MHz Atmel [23] processor with 8 kB of programming memory, and 512 B of data memory. The node is equipped with a 916 MHz, single channel, low power radio from RFM [24], capable of delivering 10 kbps of raw bandwidth using on off keying (OOK) modulation. The transmission power of the radio is dynamically tunable with different potentiometer (*pot*) settings as shown in Figure 3. For the rest of this paper, we use the mapping in Table 1 to refer to the potentiometer setting.

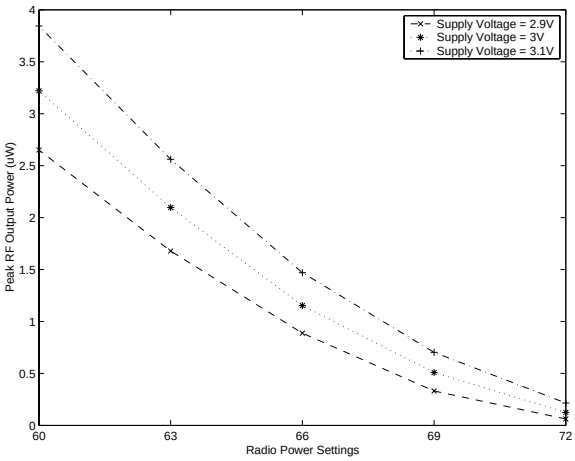


Figure 3: RF output power at different power settings.

Potentiometer Setting	Legend
60	Very High
63	High
66	Medium
69	Low
72	Very Low

Table 1: Mapping between hardware potentiometer setting and legend

To avoid undesirable effects on transmission power due to unregulated voltage supply from the batteries, all nodes

Figure 4: The Rene Mote: a hardware platform for wireless networking

are equipped with fresh AA batteries before conducting all experiments. Furthermore, all nodes have the same antenna length with uniform vertical orientation. We should, however, note that perfect calibration of the radio hardware on our experimental nodes is difficult and thus even with the same nominal hardware settings, actual transmit power on different nodes can vary [38].

The TinyOS [25] platform provides the essential runtime system support. It includes a complete network stack with bit-level forward error correction, 16-bit CRC error checking, medium access control, network messaging layer, non-volatile storage, and timing capability. The default packet size is 38 bytes long, with a payload of 30 bytes. The medium access control protocol [30] is a variant of the simple carrier sense multiple access (CSMA) protocol [39]. It waits a random duration before each transmission and goes into random backoff if the channel is busy. The delay and backoff durations are randomly picked from a fixed interval between 6ms and 100ms. During backoff, the radio is actually powered off to save energy, but the tradeoff is that no communication is possible during that period. Unlike many MAC protocols such as IEEE 802.11 [41] which will drop packet transmission after a maximum number of backoffs, this MAC protocol keeps trying until it finds a clear channel.

5 Description of Experiments & Methodology

To understand the dynamics of epidemic algorithms, we conducted two separate sets of experiments. The first

set focused on understanding the characteristics of links among all nodes in a large test bed. The second set focused on studying the dynamics of flooding over a similar test bed. Table 2 summarizes these two sets of experiments.

5.1 Experiment Set 1

For these experiments, 185 nodes were laid out over an open parking structure in a regular grid, with a grid spacing of 2 feet. The goal was to map the connectivity characteristics between all nodes at 16 different radio transmit power settings, in which nodes transmitted in sequence in response to commands sent by a base-station. The base-station issued commands to all nodes to control the experiment periodically and ensured that only one node would transmit at a time to eliminate collisions. For each transmit power setting, each node transmitted twenty packets, one node at a time. All packets were sent in sequence, 100ms apart. Receivers logged the transmitter's ID, sequence number, and transmit power setting, which were embedded in the packet payload, into their local EEPROM storage.

Prior to the experiment, nodes were subject to a diagnostic test to detect unresponsive and failed nodes, and broken antennas. Sixteen of the nodes were removed, bringing the number down to 169, which was arranged in a 13×13 grid. A total of about 54000 ($20 \times 16 \times 169$) messages were transmitted in the system, allowing us to construct a map of packet loss statistics at each power level. Some of these power levels(pot setting < 60) were beyond the useful extent of the map, and results from these settings are omitted from the analysis. This entire set of experiments was conducted over a four-hour period.

5.2 Experiment Set 2

The second set of experiments involved 156 nodes over an open parking structure, under identical settings as the first. No obstacles were present in the immediate vicinity. The nodes were laid out in a 13×12 grid, again with a 2 ft. separation. The base-station was placed in the middle of the base of the grid. The base-station initiated flooding periodically, with the period long enough to let the flood settle down. Each node would rebroadcast a message only once upon first reception of a new flood. Eight different transmit power settings were chosen and 10 non-overlapping floods were issued at each setting.

Both the application and MAC layers logged necessary information to reconstruct the epidemic message propagation. At the application layer, the identifier of the node from which a message had been received was logged. Since we used globally unique identifiers for each node, this gave us a causal ordering of message propagation [9], which was used to reconstruct the propagation tree. At the MAC layer, timing information was crucial for us to extract metrics such as backoff time and collisions. While absolute time-synchronization [26] was an option, this proved to be unnecessary for our needs. To obtain timing information, the MAC layer stored two locally generated timestamps, with granularity 16 μs. The first timestamp recorded the total amount of time that a message was stored on a node before being retransmitted. The second timestamp recorded the interval for which the node was in backoff mode. The fact that flood propagation through a large network occurs quite quickly is our ally, since clock skew and drift is small during the flooding period. However, we still had to contend with receiver delay (as noted in [26]), which we reduced to a minimum by recording timestamps at the link layer. Thus, we restricted reconstruction errors to under a bit-time per hop, which is 100 μs at 10 $kbps$.

5.3 Analysis of Experiments

As we saw in Section 2, epidemic propagation exhibits complex behaviors at scale, although the algorithm that runs on each wireless node is quite simple. Our methodology in analyzing the vast quantity of data collected during the experiments is to decompose the behavior into layers, analyze them independently with different metrics, and combine the analysis as a composite to explain the global behavior.

At the physical and link layers, we attempt to quantitatively define and measure the effective communication radius at a given transmit power in a real setting. We explore packet loss statistics over distance, define what constitutes a bidirectional link and an asymmetric link, and measure these effects. At the medium access control level, we instrument the MAC and the link layer to shed light on the degree of contention, collisions, and hidden-terminal effect that occur during the process of flooding. At the network and application layer, we analyze the resulting structure of the flood. As a composite of this analysis, we reconstruct the process of the epidemic message propagation and explain how the interactions across levels lead to the final global behavior. A comprehensive understanding of the characteristics of the radio, effective communication range, packet loss behavior, extent of asymmetry, and MAC layer behavior provides guidance for algorithms designed to work under similar large

Experiment Set	Network Size	Number of Transmit Power Settings	Comments
1	169	16	Packet loss statistics at different power levels over a grid
2	150	8	Epidemic flooding at different power levels over a grid

Table 2: Summary of the two sets of experiment

scale multihop wireless networks.

6 Physical and Link Layer Analysis

Our first step towards analyzing the data is to develop a set of metrics that will help us understand the basic link characteristics of the testbed. These metrics include node-to-node packet loss rate with respect to distance, cell radius coverage with respect to different radio transmission power setting, and the degree of link asymmetry.

6.1 Packet Loss Statistics

A fundamental metric while evaluating link-layer connectivity is packet loss. In our experiments, packets that fail to pass CRC checking are considered lost. The distribution of packet lost over distance is quite non-uniform, as Figure 5 shows. Our experimental data set, which we are making available to the research community [url - not filled in for blind review], includes detailed packet loss statistics. For instance, there are approximately 1200 packet loss data points corresponding to inter-node distances between 2 to 4 ft. We believe that these experimental statistics will be useful to drive large scale simulation studies involving such devices (for instance [7]).

Range is often described in terms of radio signal strength, which falls off polynomially but may be affected by many environmental factors. Algorithmically, what matters is successful communication. At the signal strength level, it is well known that signal propagation on the ground is $1/r^{\alpha}$ where $\alpha > 2$. However, as Figure 6 shows, the decay of packet loss with respect to distance does not experience such a sharp falloff, particularly for larger transmit power settings.

Another observation from these curves is that the throughput is lower than 100% even at short distances from the transmitter. This is due to two factors: increased fading rate due to deployment on the ground [42], and insufficient signal processing and forward error correction due to the limited computational and energy resources available on this platform.

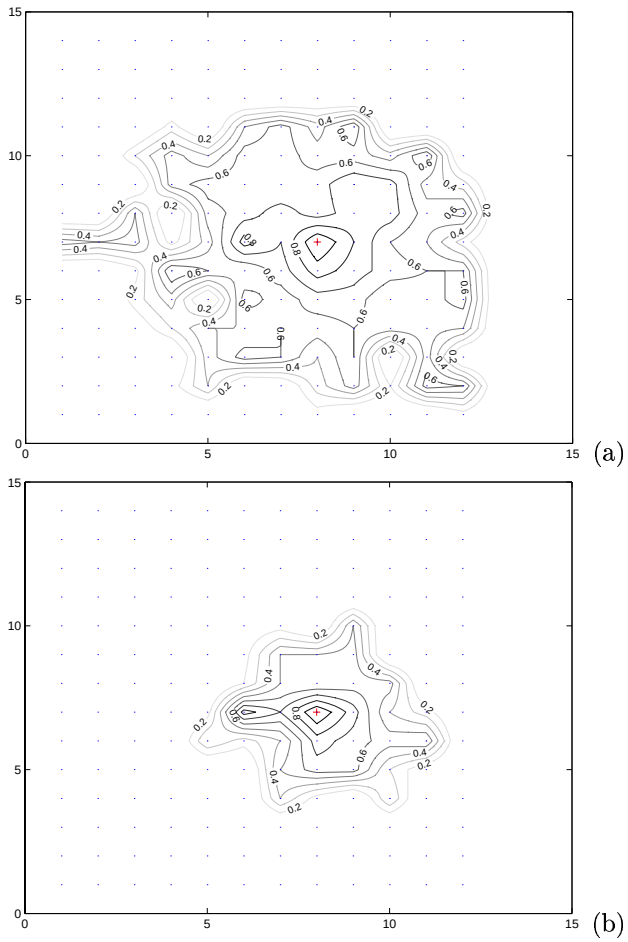


Figure 5: Contour of packet loss rate from a central node at two different transmit power settings

6.2 Measuring the Connectivity Radius

Algorithm designers often conceptualize these systems in terms of the connectivity radius and the notion of a circular connectivity cell. Many analytical results involve working with circular cells, since this simplifies analysis and allows a geometric approach. Figure 5 shows that connectivity is not a simple binary relation on distance.

Figure 6 shows that packet loss decreases monotonically with distance. The definition of connectivity radius is

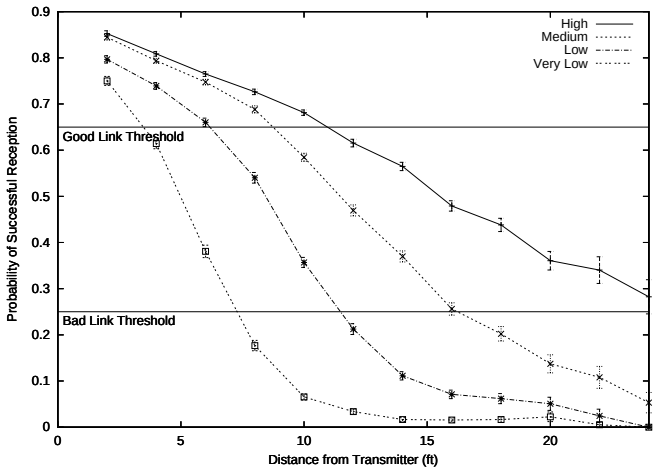


Figure 6: Probability of packet reception over distance with different transmission power settings.

typically based on a packet-loss threshold. A guiding argument for choosing this threshold would be to treat a link as a "good link" if we can use forward error correction(FEC) and other techniques to improve the raw packet throughput to adequate levels. Correspondingly, a "bad link" would be regarded as one which cannot possibly be salvaged by such means, as it offers very poor throughput. Using these criteria, and based on Figure 6 we take the threshold for a "good link" to be 65% and the corresponding threshold for a bad link to be 25%[2].

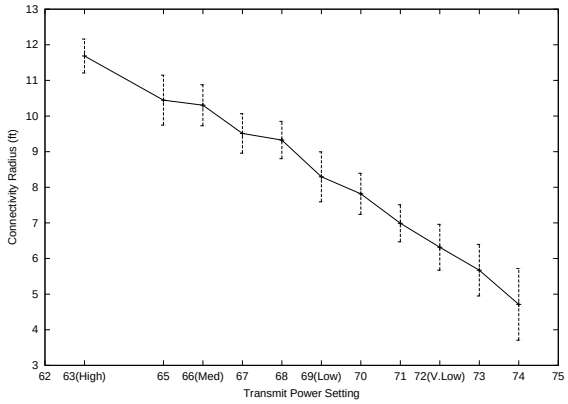


Figure 7: Communication radius at different radio transmission power.

We define the *connectivity radius* of a node as the radius R of the smallest circle that covers 75% of nodes that have greater than 65% throughput to n. Only 16 nodes in the center of the grid were considered for this calculation, since their cells completely fit within the grid for most transmit power settings studied. Figure 7 shows that there is a linear variation of the connectivity radius with the transmit power setting on the mote.

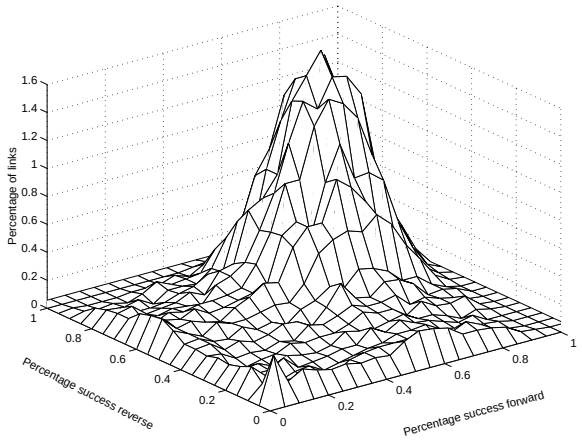


Figure 8: Histogram on the probability of reception in the forward versus reverse direction for all pairs of nodes at low transmit power setting.

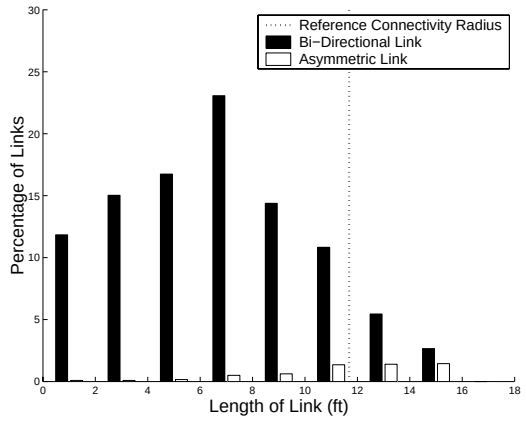
6.3 Asymmetric Links

Asymmetric links arise relatively infrequently in sparse wireless networks, such as typical 802.11 LAN and ad-hoc configurations, and are often filtered out by protocol levels [1, 37]. However, with a large field of low-power wireless nodes such asymmetric links are very common, even if all nodes are set to have the same transmit power. Our experiments allow us to quantify asymmetric links and understand their behavior. The definitions used for a "good" and "bad" link in Section 6.2 is used in developing this metric as well. An *asymmetric link* is defined as one which has a "good" link in one direction and a "bad" link in the other. A *bidirectional link* is one which has a good link in both directions. The distribution of asymmetric links over the entire network is shown in Figure 8. This analysis reveals that for the range of transmit power settings studied, approximately 5-15% of all links are asymmetric, the percentage increasing with decreasing transmit power setting.

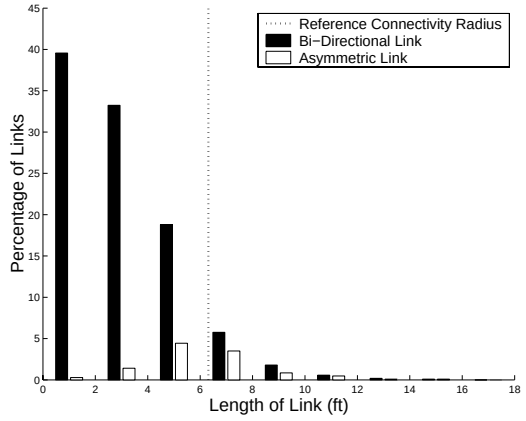
Figure 9 shows the distribution of bi-directional and asymmetric links over distance. At short distances from the transmitter, a negligible percentage of links are asymmetric, but this percentage grows significantly with increasing distance, especially at lower power settings. The dotted vertical line shows the connectivity radius calculated as per section 6.2 for the particular transmit power.

At the fading edge of a connectivity cell, small differences

[2]Links that lie between somewhere in between these could be potentially good if the probability of successful reception improved over time. We do not consider these links further, although they would be significant to a more long-term empirical study.



High transmit power



Very low transmit power

Figure 9: Distribution of bidirectional and asymmetric links over distance.

between the nodes in transmit power and reception sensitivity become significant, resulting in asymmetry. The aggregate effect of the small differences in the radios and hardware, as mentioned in Section 4, and slight differences in energy levels of the nodes contribute significantly to link asymmetries in this regime (discussed further in Section 9).

7 Medium Access Layer Analysis

We now turn to the dynamics and MAC layer effects during message propagation. We examine four metrics that capture different aspects of the propagation: maximum backoff interval, reception latency, settling time and useless broadcasts[3].

[3]We use 95% thresholds to meaningfully capture the effect of transmit power setting, interference cell and edge effects on MAC layer behavior.

7.1 Maximum Backoff Interval

The distribution of backoff intervals in the network indicates the extent of contention that each node perceives in the channel. In analyzing contention, it is important to consider the interference range, which is often greater than the communication range. As the transmit power setting increases, contention is greater since the interference cell grows larger. In our experiment, this result in nodes being forced to back off for longer durations at higher transmit power settings, as shown in Figure 10.

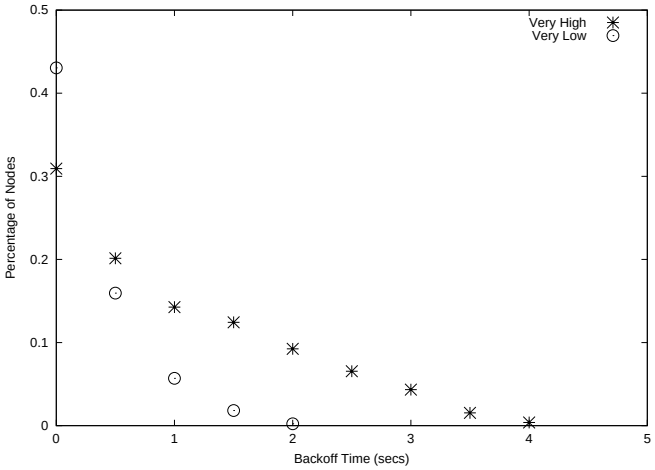


Figure 10: Distribution of backoff delay before successful transmission at two transmission power settings.

A metric that captures the contention level within an interference cell is the *maximum backoff interval* among all nodes. This reflects the time till contention subsides in each cell. The measure, however, is approximate, since different nodes within the cell may receive the packet at different times, and might not initiate backoff simultaneously. Table 3 shows that the transmit power setting and the 95% maximum backoff interval were directly proportional to one another, over the range of transmit power settings studied.

7.2 Reception Latency

We now turn to evaluating the *Reception Latency*, which we define as the amount of time for nodes in the network to receive an epidemic broadcast packet. Figure 11 shows three time-series plots for different transmit power settings. An interesting observation is that a significant fraction of the total epidemic propagation time was taken to reach the last few nodes in each plot. These last few nodes captured by the propagation are significant in the application context, and form stragglers that were captured

Power Setting	#Expts	95% MaxBackoff Interval (s)	95% Reception Latency (s)	Network Diameter (hops)	95% Settling Time (s)	% Useless Broadcasts(%)
Very High	7	3.265 ± 0.169	1.285 ± 0.302	5	3.842 ± 0.330	82.2 ± 2.3
High	8	2.773 ± 0.099	1.569 ± 0.278	5.875 ± 0.295	3.663 ± 0.125	78.4 ± 2.2
Medium	7	2.587 ± 0.075	1.688 ± 0.307	6.286 ± 0.451	3.469 ± 0.172	75.0 ± 3.8
Low	4	2.202 ± 0.073	1.861 ± 0.476	7 ± 1.299	3.258 ± 0.185	70.4 ± 6.6
Very Low	4	1.302 ± 0.057	2.174 ± 0.235	9	2.985 ± 0.161	63.6 ± 3.3

Table 3: Maximum backoff interval, reception latency, settling time, useless broadcasts, with corresponding 95% confidence intervals at different transmit power settings.

by the propagation on its rebound. We expand upon this observation in section 7.5. Table 3 and Figure 11 also show the relationship between reception latency and the *network diameter*, which refers to the maximum number of hops from the source to any node in the network. As expected, for higher transmit powers, the reception latency decreased with the network diameter. The shape of the curve in Figure 11 reflects the mechanics of flood propagation. The number of nodes covered starts slowly as the flood picks up more nodes, then grows rapidly as many of these nodes retransmit in rapid succession, and tapers out slowly as the few remaining nodes are picked up. An interesting observation is that a significant fraction of the total epidemic propagation time was taken to reach the last few nodes in each plot.

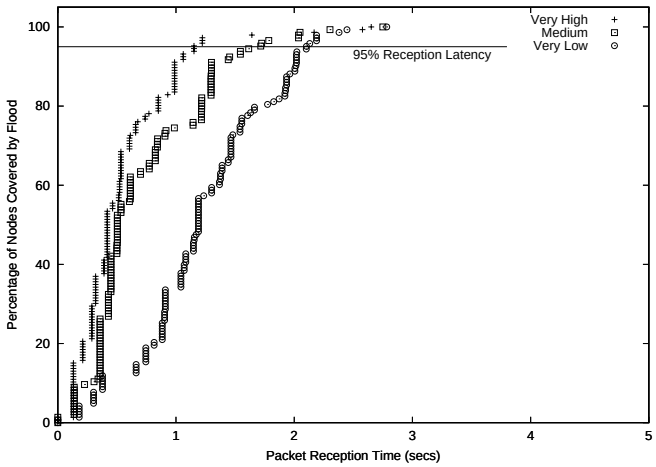


Figure 11: Latency grows with Increasing Transmit Power Setting.

7.3 Settling Time

We define *Settling Time* as the time taken for delivery of a single packet flood throughout the network. This is a combination of the time taken for the flood to propagate to the far end of the network(reception latency), and for all retransmissions to complete within each communication cell (maximum backoff interval). The three metrics can be related as follows:

$$max(MaxBackoffTime, ReceptionLatency) \leq Settling\ Time \leq MaxBackoffTime + ReceptionLatency$$

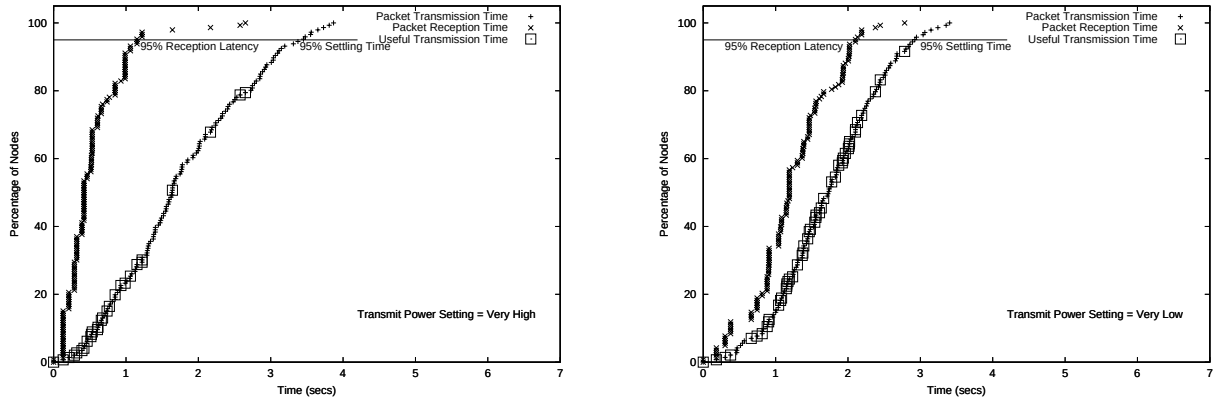
The above relation has a simple explanation: the settling time is at least as long as the time until the node with the longest backoff time has retransmitted a packet or the last node receives a packet. It is bounded above by the case that the last node to receive the epidemic propagation also chooses the maximum backoff interval.

Table 3 sheds light on the relation between these metrics and the transmit power level. At low transmit power settings, the settling time is closer to the reception latency than the maximum backoff interval, suggesting that the end-to-end flood propagation delay has a more significant impact than time taken for broadcasts to subside within each interference cell. This follows intuition since the network diameter is larger at low transmit power settings. As the transmit power increases, the settling time is closer to the maximum backoff time; the network diameter is low, and the contention within each cell dominates. The relationship between settling time and reception latency is also captured in Figures 12(a) and 12(b), which shows the time-series of the propagation of the flood. At high transmit power setting, reception latency is a small fraction of the settling time.

The difference between the settling time and reception latency in Table 3 also shows that at high transmit power settings, nodes in the network keep retransmitting the message long after 95% of the network has received the packet. We now look at such rebroadcasts, that do not reach additional nodes.

7.4 Useless Broadcasts

We define *useless broadcasts* as the percentage of rebroadcasts that deliver a message only to nodes that have already received one. These can occur because all neighbors



(a) High transmit power setting.

(b) Very Low transmit power setting.

Figure 12: Timeseries of packet transmission and reception at different transmit power settings.

have already received the message, or the rebroadcast suffers packet loss or collision. The time-series in Figure 12(a), shows that for high transmit power, a significant fraction of the broadcasts are useless since most nodes are covered by the flood quite early. In contrast, Figure 12(b) shows that each additional transmission is useful towards capturing nodes by the flood when the transmit power setting is low.

Table 3 shows the percentage of useless broadcasts over a range of transmit power settings. While the "Very Low" transmit power setting has around 60% useless rebroadcasts, we observe that reducing the transmit power setting further significantly decreases the probability of reaching all nodes in the network.

7.5 Collision

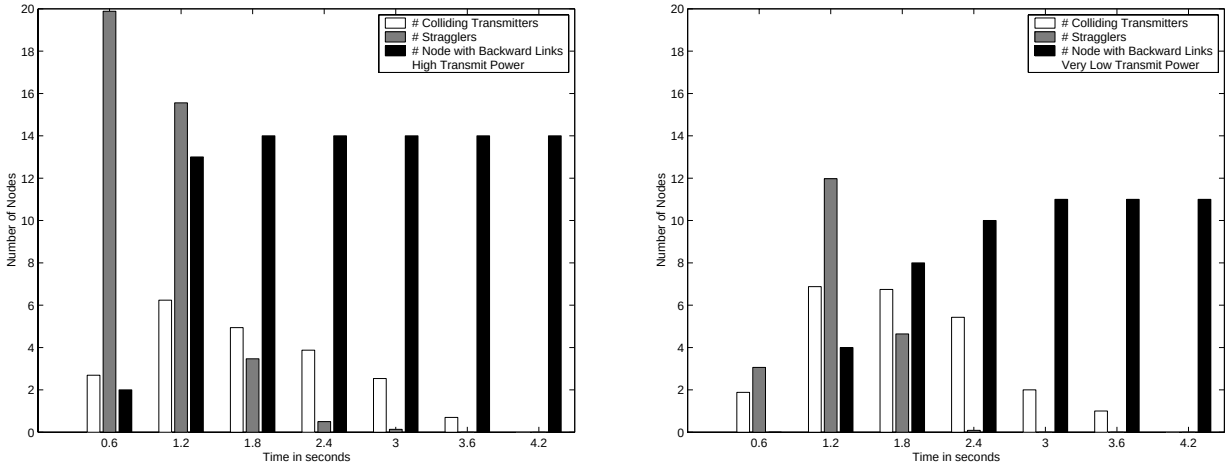
We notice in Figures 11 and 12, that the last 5% is significantly different from the rest. In fact, Figure 11 shows that for all plots, the time taken for all the nodes(100%) to receive the flood is almost equal, although this hardly reflects the significant differences between the curves. Similarly, in Figure 12(b), the last 5% of the nodes take as much time to receive their packets as the first 95%. These observations relate to stragglers and backward links, introduced in Section 2. We define *stragglers* as nodes that miss all transmissions, even though they would be expected to receive a packet with high probability. *Backward links* are defined as links in which the recipient of the flood is closer from the base station than the transmitter. In this section, we understand how collisions can explain some of this behavior.

As discussed in Section 4, control packets such as RTS/CTS are often used to deal with the hidden-terminal problem [40]. In broadcast-style epidemic transmission, a packet does not have an intended recipient, so CSMA without RTS/CTS is used. The existence of signal interference across cells and link asymmetry complicate the collision effect, making it difficult to detect and measure. However, we are able to combine the global ordering of message transmission and link layer estimates of the communication cell to infer the impact of collisions.

Figure 13 shows the relation between number of colliding transmitters, stragglers and backward links. Medium access layer timestamps are used to estimate the number of colliding nodes i.e. nodes whose transmissions overlap. Using the connectivity radius in Figure 7, we look at intersecting communication cells for each set of colliding transmitters. This gives us an estimate of the number of nodes that should receive the packet, but do not as a result of collisions. These stragglers miss the propagation in the early stage of the flood and form backward links from a later reception. At higher transmit power setting, each node has a larger communication cell, and the number of hidden terminals (reflected by stragglers) is larger. This results in a larger number of backward links being generated, as the flood rebounds to capture stragglers.

8 Network and Application Layer Analysis

We now look at a tree construction based on the simple epidemic algorithm to give us an idea of the significance



(a) High transmit power setting.

(b) Very Low transmit power setting.

Figure 13: Histogram showing distribution of colliding transmitters, distribution of stragglers and cumulative distribution of backward links over time

of our measurements in an application context. Consider a routing tree that is constructed as the reverse path of the epidemic propagation. These trees are constructed as multi-hop reverse paths from every node back to the initial source and are useful in data-gathering applications such as wireless sensor networks [43]. The reverse paths setup during flooding is also a step in most reactive mobile ad-hoc routing algorithms with caching [10], and multicast algorithms [17, 18].

Our link layer measurements have interesting implications. Long links have a greater potential for being asymmetric and are therefore less appropriate for use as the reverse path to draw data back to the source. It should be noted that an asymmetric long link closer to the root will orphan larger sub-trees. Tree reconfiguration is expensive, and should be avoided by filtering asymmetric links during epidemic propagation. Similarly, backward links in the tree are sub-optimal, since data flows away from the base-station, rather than towards it.

Figure 14 shows a combination of factors affecting the tree structure. Here, level represents the number of hops between the base-station and a node on the tree, and the distance corresponds to the physical distance between them. At one end (upper left), nodes that were physically close to the base-station were many more hops away from it than their peers.This can be explained by the existence of stragglers. In the lower right, nodes far away from the base-station were few hops from it due to long links.

The parent selection mechanism has a large role in influencing the structure of the tree that is built. A simple parent selection mechanism is for each receiver to select the source of the first message it received as its parent. This opportunistic, earliest-first, parent selection mechanism results in highly clustered trees, with most nodes being leaf nodes and yet a significant number having large number of children. We examine the behavior of this mechanism by reconstructing the trees based on our data logs from the experiments. Figure 15 shows histograms of the cluster sizes for this parent selection mechanism for two transmit power settings, across various runs. The loglog plots indicate that this distribution has a relatively heavy tail, with large cluster sizes occurring frequently. This behavior was observed across the different power settings being studied. As we reduced the power setting, the slope of the least squares fit increased, as a result

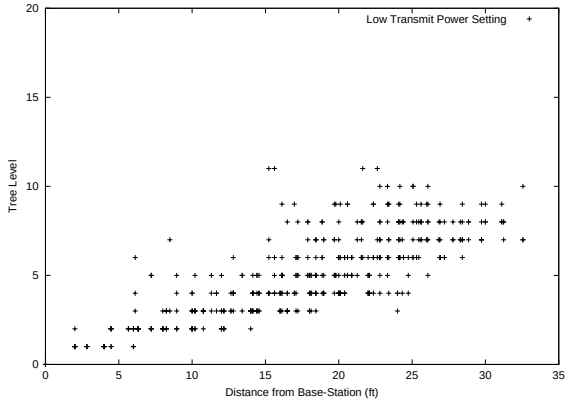


Figure 14: Level of the nodes in the constructed tree against node distance from the root of the tree.

of decreasing communication cell size. High clustering is exacerbated by the presence of long links. Nodes at the end of long links have a greater possibility of seeing less interference and have many neighbors who have not been covered in the epidemic propagation. These nodes, therefore, retransmit the packet faster, and reach more uncovered nodes, resulting in the high clustering behavior. However, longer links are likely to be asymmetric and hence not suitable for reverse-path routing.

9 Discussion and Conclusion

We have identified the complexities in a simple epidemic algorithm by examining contributions from the various layers. Recall the scenario presented in Section 2 where we discussed four notable effects: long links, backward links, stragglers and clustering. The incidence of long links is explained in our physical/link-layer discussion (Section 6.1). Figure 6 shows that while packet loss increases with distance, it has a fairly long tail. With many nodes, it is likely that a given transmission will reach some nodes far away. Stragglers can be explained by collision effects caused at the MAC layer (Section 7.5). Backward links can be explained as a combination of the effect of long links and collisions. Long links resulted in the epidemic propagating faster in certain directions, and rebounding to fill areas where the propagation was slower or where stragglers remained, forming backward links. This opportunistic, earliest-first, parent selection mechanism at the network/application layer, results in highly clustered trees (Section 8).

In analyzing the contributions of various layers, we have defined and studied useful metrics at each level. We now point out implications of our results concerning these metrics and pose some open questions that, we hope, can be answered with further research.

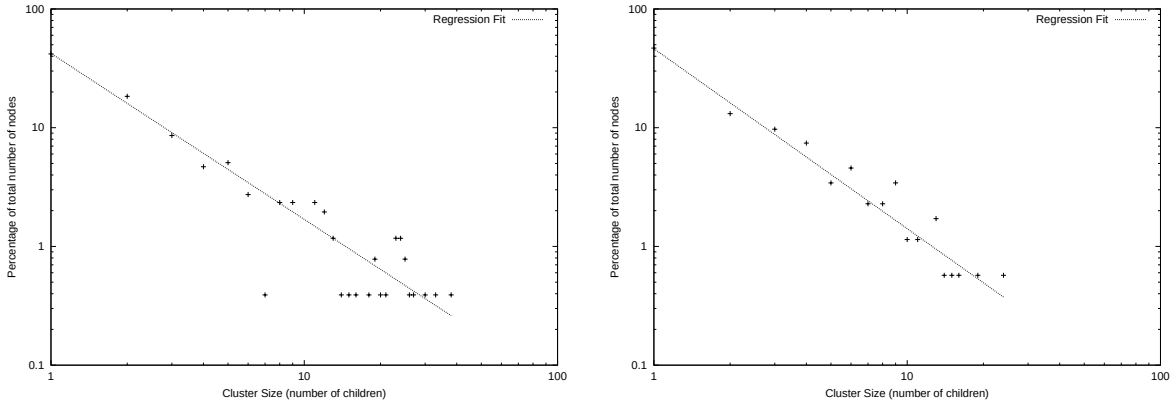
Much of the literature has assumed circular disc model for the cell regions. Our data implies that a probabilistic view of modelling neighborhood is more reasonable. For a deterministic abstraction of this probabilistic approach, the medium access or topology construction layer may provide algorithms with a neighbor abstraction that include only a subset of nodes that are consistently in range. Yet, providing such an interface incurs significant overhead, which may be hard to justify for highly resource constrained nodes. MAC layer protocols such as 802.11 incur processing overhead (e.g. synchronization beacons) to provide such a deterministic abstraction. Our data can be used to drive large-scale simulations to validate algorithm behavior over a probabilistic model of the underlying node connectivity.

Other observations at the link layer are equally significant. The presence of long links might have unanticipated manifestations at scale. For trees constructed using epidemic algorithms, long links should be avoided as they tend to be asymmetric. The fact that asymmetry manifests itself more on long links (Figure 9) is significant to many ad-hoc routing protocols that use shortest reverse hop-count as a method for setting up routing paths. These protocols *naturally select* links that are long in nature, and thereby increase the probability of selecting a poor reverse end-to-end route.

Some of the ad hoc routing protocols proposed in the literature, such as DSR [10] and ZRP [11], can route using asymmetric links; however there are others, such as AODV [12] and TORA [13], that either assume that all links are symmetric or filter out asymmetric links. Our experimental study suggests that asymmetric links are indeed likely to be significant in large scale, multihop networks and that robust protocols must deal appropriately with asymmetric links through mechanisms such as the use of a sub-routing layer to provide bidirectional abstraction [37].

Our analysis of the medium access layer shows how to apply simple techniques to obtain fine-grained timing information in a large distributed network. Our data also provide useful hints to augment the MAC layer for epidemic message propagation. Reducing the energy consumption of the epidemic broadcast can be served by using low transmit power, and reducing the number of useless broadcasts. Section 7.4 discusses the amount of redundancy that is necessary if one wishes to reliably propagate the flood at low transmit power. Rebroadcasts that are transmitted after a large backoff delay relative to the elapsed time of the flood are likely to be useless. Therefore, energy can be saved by dropping these rebroadcasts if tight estimates of the reception latency are available to nodes.

Another way to augment the performance of the epidemic algorithm is to increase its throughput. The settling time metric that we defined in Section 7.3 is useful to describe *Multicast Throughput* as the number of packets per second that can be pumped into the wireless network. In order to maximize throughput, it is useful to pipeline transmissions from the base-station, such that two successive transmissions are separated by the amount of time it takes to propagate two transmission cells [30]. Settling time gives a lower bound on this throughput, by quantifying the time taken for the entire network to settle down. Thus a lower bound on achievable multicast throughput is 1/settling time.



(a) High Transmit Power Setting

(b) Low transmit power setting.

Figure 15: Cluster size histogram in log log scale.

In conclusion, our study has implications for several algorithm design and measurement considerations:

- Simple protocols with very few states can exhibit unanticipated global complexity due to their interaction with the complex physical world.
- Vertically integrated measurement is useful to understand the contribution of the physical, medium access and application layer to application behavior in scale.
- Algorithm designs should use a probabilistic abstraction to model connectivity.
- Asymmetry is to be expected, and certain protocol choices may exacerbate its effect. Robustness to asymmetry is a crucial part of protocol design in these systems.

References

[1] D-K. Kim, C-K. Toh, and Y-H. Choi, "On Supporting Link Asymmetry in Mobile Ad Hoc Networks," *Proceedings of IEEE GLOBECOM 2001*, San Antonio, Texas, November 2001.

[2] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, J.-P. Sheu, "The Broadcast Storm Problem in a Mobile Ad Hoc Network," *Mobicom '99*, Seattle, Washington, USA.

[3] D.A. Maltz, J. Broch, D.B. Johnson, "Experiences Designing and Building a Multi-Hop Wireless Ad Hoc Network Testbed," *CMU School of Computer Science Technical Report CMU-CS-99-116*. March 1999.

[4] S. Desilva and S.R. Das, "Experimental Evaluation of a Wireless Ad Hoc Network," *Proceedings of the 9th Int. Conf. on Computer Communications and Networks (IC3N)*, Las Vegas, October 2000.

[5] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart and D. Terry, "Epidemic Algorithms for Replicated Database Maintenance," In *Proceedings of the Sixth Symposium on Principles of Distributed Computing*, pages 1-12, August 1987

[6] J. M. Kahn, R. H. Katz and K. S. J. Pister, "Mobile Networking for Smart Dust," *ACM/IEEE Intl. Conf. on Mobile Computing and Networking (MobiCom 99)*, Seattle, WA, August 17-19, 1999

[7] Phil Levis. "TinyOS simulator," *http://tinyos.net*

[8] Harold Abelson et al, "Amorphous Computing," *Communications of the ACM*, Volume 43, Number 5, May 2001.

[9] Leslie Lamport, "Time, Clocks, and the Ordering of Events in a Distributed System," *Communications of the ACM* vol 21, no 7, p. 558-565, 1978

[10] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networking," *Mobile Computing*, T. Imielinski and H. Korth, Eds., Kluwer, 1996.

[11] M. R. Pearlman and Z. J. Haas, "Determining the Optimal Configuration for the Zone Routing Protocol," *IEEE Journal on Selected Areas in Communications: Wireless Ad-Hoc Networks*, vol. 17, no. 8, p. 1395-1414, August 1999.

[12] C. E. Perkins and E. M. Royer, "Ad hoc on-demand distance vector routing," *IEEE WMCSA*, vol. 3, p. 90-100, 1999.

[13] V. Park and M.S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," Proc. IEEE INFOCOM '97, Kobe, Japan, Apr 1997.

[14] B. Das and V. Bharghavan, "Routing in ad-hoc networks using minimum connected dominating sets," *IEEE Internation Conference on Communications*, p. 376-380, June 1997.

[15] K.M. Alzoubi, P.-J. Wan, O. Frieder, "New Distributed Algorithm for Connected Dominating Set in Wireleess Ad Hoc Networks," *Proceedings of the 35th Hawaii International Conference on System Sciences*, Big Island, Hawaii, 2002.

[16] J. Wu, and H. Li, "Domination and its Applications in Ad Hoc Wireless Networks with Unidirectional Links," *Proc. of International Conference on Parallel Processing (ICPP)* Aug. 2000, 189-200.

[17] C. Ho, K. Obraczka, G. Tsudik, K. Viswanath, "Flooding for Reliable Multicast in Multi-Hop Ad Hoc Networks," *ACM DIAL M '99*, Seattle, Washington, USA, 1999.

[18] M. Gerla, C.-C. Chiang, and L. Zhang, "Tree Multicast Strategies in Mobile, Multihop Wireless Networks," *ACM/Baltzer Mobile Networks and Applications Journal*x, 1998

[19] E. Pagani, G.P. Rossi, "Reliable Broadcast in Mobile Multihop Packet Networks". *Proceedings 3rd ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM'97)*, Budapest, 26-30 Sep. 1997, pp. 34-42.

[20] R. Chandra, V. Ramasubramanian, and K. P. Birman, "Anonymous Gossip: Improving Multicast Reliability in Mobile Ad-Hoc Networks", International Conference on Distributed Computing Systems, 2001.

[21] L. Li, J. Halpern, Z. J. Haas, "Gossip-based Ad Hoc Routing," unpublished.

[22] D. Ganesan *et al.* "Large-scale Network Discovery: Design Tradeoffs in Wireless Sensor Systems," poster presented at the 18th ACM Symposium on Operating System Principles, Banff, Canada, October 2001. *http://lecs.cs.ucla.edu/ẽstrin/.*

[23] ATMEL 8-bit RISC Processor *http://www.atmel.com/atmel/products/prod23.htm*

[24] RF Monolithics *http://www.rfm.com/products/data/tr1000.pdf*

[25] J. Hill *et al.*, "System architecture directions for network sensors," in *ASPLOS* 2000

[26] J. Elson and D. Estrin, "Time Synchronization for Wireless Sensor Networks," *Proceedings of the 2001 International Parallel and Distributed Processing Symposium (IPDPS), Workshop on Parallel and Distributed Computing Issues in Wireless and Mobile Computing*, San Francisco, California, USA. April 2001.

[27] Jeremy Elson, Lewis Girod and Deborah Estrin, "Fine-Grained Network Time Synchronization using Reference Broadcasts". *Submitted for review*, February 2002.

[28] J. Heidemann *et al.* "Building Efficient Wireless Sensor Networks with Low-Level Naming," in *Proceedings of the Symposium on Operating Systems Principles*, pp. 146-159. Banff, Alberta, Canada, October, 2001

[29] W. Ye, J. Heidemann, and D. Estrin, " An Energy-Efficient MAC Protocol for Wireless Sensor Networks, IEEE INFOCOM, New York, NY, USA, June, 2002.

[30] A. Woo and D. Culler, "A Transmission Control Scheme for Media Access in Sensor Networks," *Mobicom 2001.*

[31] S. Klemmer, S. Waterson, and K. Whitehouse, "Towards a location-based context-aware sensor infrastructure," *unpublished*, 2000. http://guir.berkeley.edu/projects/location/Location.pdf

[32] J. Broch *et al.* A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols *Mobicom '98*, ACM, Dallas, TX, October 1998.

[33] B. Krishnamachari, D. Estrin, and S. Wicker, "Impact of Data Aggregation in Wireless Sensor Networks," *DEBS'02.*

[34] D. Estrin *et al.*"Next Century Challenges: Scalable Coordination in Sensor Networks," *ACM/IEEE International Conference on Mobile Computing and Networks (MobiCom '99)*, Seattle, Washington, August 1999.

[35] C. Intanagonwiwat *et al.*, "Impact of network density on data aggregation in wireless sensor networks," *submitted for publication to International Conference on Distributed Computing Systems (ICDCS-22)*, November 2001.

[36] W. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," *Proc. 5th ACM/IEEE Mobicom Conference (MobiCom '99)*, Seattle, WA, August, 1999.

[37] V. Ramasubramanian, R. Chandra and D. Mosse, "SRL: A Bidirectional Abstraction for Unidirectional Ad-Hoc Networks," *INFOCOM 2002*, June 23-27, New York, 2002.

[38] J. Hightower, C. Vakili, G. Borriello, and R. Want, "Design and Calibration of the SpotON Ad-Hoc Location Sensing System", *unpublished*, August 2001.

[39] S.S. Lam. "A carrier sense multiple access protocol for local networks," *In Computer Networks*, volume 4, pages 21-32, 1980.

[40] D. Allen. "Hidden terminal problems in Wireless LAN's," IEEE 802.11 Working Group Papers, 1993.

[41] ANSI/IEEE Std 802.11 1999 Edition.

[42] K. Sohrabi, B. Manriquez, and G. Pottie, "Near Ground Wideband Channel Measurement", *Vehicular Technology Conference IEEE*, volume 1, pages 571-574, 1999.

[43] C. Intanagonwiwat, R. Govindan and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," *ACM/IEEE International Conference on Mobile Computing and Networks (MobiCom 2000)*, August 2000, Boston, Massachusetts